



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2006

Empathic neural responses are modulated by the perceived fairness of others

Singer, T ; Seymour, B ; O'Doherty, J P ; Stephan, K E ; Dolan, R J ; Frith, C D

Abstract: The neural processes underlying empathy are a subject of intense interest within the social neurosciences. However, very little is known about how brain empathic responses are modulated by the affective link between individuals. We show here that empathic responses are modulated by learned preferences, a result consistent with economic models of social preferences. We engaged male and female volunteers in an economic game, in which two confederates played fairly or unfairly, and then measured brain activity with functional magnetic resonance imaging while these same volunteers observed the confederates receiving pain. Both sexes exhibited empathy-related activation in pain-related brain areas (fronto-insular and anterior cingulate cortices) towards fair players. However, these empathy-related responses were significantly reduced in males when observing an unfair person receiving pain. This effect was accompanied by increased activation in reward-related areas, correlated with an expressed desire for revenge. We conclude that in men (at least) empathic responses are shaped by valuation of other people's social behaviour, such that they empathize with fair opponents while favouring the physical punishment of unfair opponents, a finding that echoes recent evidence for altruistic punishment.

DOI: <https://doi.org/10.1038/nature04271>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-50407>

Journal Article

Accepted Version

Originally published at:

Singer, T; Seymour, B; O'Doherty, J P; Stephan, K E; Dolan, R J; Frith, C D (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature*, 439(7075):466-469.

DOI: <https://doi.org/10.1038/nature04271>

Published in final edited form as:

Nature. 2006 January 26; 439(7075): 466–469. doi:10.1038/nature04271.

Empathic neural responses are modulated by the perceived fairness of others

Tania Singer, Ben Seymour, John P. O'Doherty, Klaas E. Stephan, Raymond J. Dolan, and Chris D. Frith

Abstract

The neural processes underlying empathy are a subject of intense interest within the social neurosciences^{1–3}. However, very little is known about how brain empathic responses are modulated by the affective link between individuals. We show here that empathic responses are modulated by learned preferences, a result consistent with economic models of social preferences^{4–7}. We engaged male and female volunteers in an economic game, in which two confederates played fairly or unfairly, and then measured brain activity with functional magnetic resonance imaging while these same volunteers observed the confederates receiving pain. Both sexes exhibited empathy-related activation in pain-related brain areas (fronto-insular and anterior cingulate cortices) towards fair players. However, these empathy-related responses were significantly reduced in males when observing an unfair person receiving pain. This effect was accompanied by increased activation in reward-related areas, correlated with an expressed desire for revenge. We conclude that in men (at least) empathic responses are shaped by valuation of other people's social behaviour, such that they empathize with fair opponents while favouring the physical punishment of unfair opponents, a finding that echoes recent evidence for altruistic punishment.

Empathy enables us to share the emotion, pain and sensation of others. The perception–action model of empathy states that the observation or imagination of another person in a particular emotional state automatically activates a representation of that state in the observer¹. Recent imaging studies provide evidence for common activation elicited when experiencing disgust⁸, touch⁹ or pain^{10–12} in oneself, and when perceiving the same feelings in others. For example, studies on empathy for pain have found that the activation observed in anterior insula/fronto-insular cortex (AI/FI) and the anterior cingulate cortex (ACC) to pain in oneself is also seen when observing pain in someone else^{10,12}. These results suggest that our ability to empathize relies on neuronal systems that underpin our own bodily and emotional states^{1,2,10,13}. However, it is unclear whether or how these responses are modulated by the social relations between individuals.

To address this question, we measured brain responses when individuals empathized with the pain of someone they liked or disliked. We used an economic game model to induce liking or disliking of two confederate actors, previously unknown to our experimental subjects. The confederates played fair or unfair strategies in a sequential Prisoner's Dilemma game (PDG) with the subjects. As illustrated in Fig. 1a, in this game the subject was always 'first mover' and could either trust the other player by sending his/her money to him/her or mistrust him/her by keeping it. The confederates were always 'second mover' and could choose between a fair or an unfair response by returning high or low amounts of money (Supplementary

Information). On the basis of previous findings¹⁴, we expected that subjects would come to like fair players and dislike unfair players.

As Fig. 1b illustrates, subjects perceived the confederates as being fair and unfair according to their game-playing strategy. Post-scan behavioural ratings confirmed that both male and female subjects rated the fair player as being significantly more fair, more agreeable, more likeable and more attractive than the unfair player (range of scale p2 to 22; P , 0.001; Supplementary Table S1).

In the second part of the experiment we used functional magnetic resonance imaging (fMRI) to investigate whether the liking or disliking acquired during the preceding game modulated empathic responses for pain. One actor sat on each side of the scanner, enabling the scanned volunteer to observe the hands of the fair and unfair player as well as his/her own hand. Painful stimulation was applied through electrodes to the hands of all three participants. As shown in Fig. 1c, cues (coloured arrows) were presented in random order indicating whether she/he (self condition), the fair player (fair condition) or the unfair player (unfair condition) would get low stimulation (no pain condition) or high stimulation (pain condition).

We predicted that pain-related empathic responses in AI/FI and ACC would be elicited when observing a fair person in pain but that this activity would be reduced or absent when observing pain in a person who had previously played unfairly. In view of recent economic models of social preferences⁴⁻⁷ and altruistic punishment¹⁵⁻¹⁷ we further expected that the reduction of empathy for an unfair person might be accompanied by an increase in activity in brain areas known to have a key function in reward processing, such as ventral striatum/nucleus accumbens and orbito-frontal cortex¹⁸⁻²⁰. It has been shown that people reward others for cooperative behaviour but punish violations of social fairness even at a personal cost^{15,16}, an effect likely to be mediated by neural mechanisms that provide intrinsic motivation (that is, satisfaction) from punishing violators²¹.

The comparison of brain activity associated with painful and non-painful trials in the 'self' condition for men and women revealed an expected increase in the 'pain network', including activity in AI and ACC (Supplementary Tables 2 and 3). Moreover, we observed pain-related empathic responses in both genders in bilateral AI extending into FI and brainstem when seeing an unfamiliar but likeable person in pain (Fig. 2a, b). The activation in ACC was significant in women ([9, 18, 27]; P , 0.001 uncorrected, P , 0.05 whole-brain corrected) but was borderline in men ([29, 39, 27]; P $1/4$ 0.001 uncorrected; Supplementary Tables 4 and 5; coordinates refer to the peak of activations in MNI (Montreal Neurological Institute) space). Furthermore, we extended previous findings¹⁰ by showing that men, as well as women, who scored higher on standard empathy scales²² had higher empathy-related brain activity in ACC and AI/FI (Supplementary Fig. 2).

To investigate whether empathic responses are modulated as a function of the perceived fairness of others we identified, for men and women, peak voxels of activation in bilateral FI observed when pain was applied to self and to fair players from the analysis described above (see Fig. 2a, b). Figure 2c, d illustrates the average activation (parameter estimates) for painful–non-painful stimulation in these voxels when subjects observed either fair or unfair players in receipt of pain. This analysis revealed that less empathic activity was elicited by the knowledge that an unfair player was in pain. However, there was also a marked difference between the sexes. In women this reduction in activity was very small, whereas in men the knowledge that an unfair player was in receipt of pain elicited no increase in empathic activity in FI. And indeed, formal analysis revealed no significant difference for women when comparing painful trials for fair versus unfair players in empathy-related pain regions. However, men showed significantly enhanced activation in bilateral FI when observing fair compared with unfair

players in pain (Supplementary Tables 8 and 11). Consistent with this finding, supplementary analysis showed that women but not men displayed significant activation in bilateral AI/FI and ACC in all three conditions (Supplementary Tables 7 and 10).

We next sought evidence of increases in brain activity in regions associated with reward processing (ventral striatum/nucleus accumbens and orbito-frontal cortex) when observing an unfair player receiving painful stimulation. As shown in Fig. 3a, we observed increased activation in left ventral striatum/nucleus accumbens for men, but not for women, in the comparison of painful trials in the unfair and the fair condition (pain in unfair–pain in fair) as well as in the interaction ((pain–no pain) unfair–(pain–no pain) fair). This latter comparison also revealed increased activation in left orbito-frontal cortex (Supplementary Table 11). As illustrated in Fig. 3b, the direct gender comparison for painful trials in the unfair and the fair condition ((pain in unfair–pain in fair) men–(pain in unfair–pain in fair) women) revealed that men showed significantly higher activation than women in left nucleus accumbens (in the same peak as identified in the former analysis).

To explore the role of nucleus accumbens and orbito-frontal cortex further, we assessed whether individual differences in an expressed desire for revenge covaried with brain activity in these regions. We derived a ‘revenge’ composite score from post-experiment questionnaires measuring a subjectively expressed desire for revenge (see Methods and Supplementary Information). Figure 3c shows that men expressed a stronger desire for revenge than women ($t(30) = 2.40$, $P = 0.05$; Supplementary Fig. 3). As illustrated in Fig. 3d, regression analysis confirmed that men, but not women, who expressed a stronger desire for revenge showed greater activation in nucleus accumbens when they perceived an unfair player receiving painful stimuli than when they perceived a fair player in pain (Supplementary Fig. 4).

Our data provide neurobiological evidence on how fairness in social interactions shapes the nature of the affective link between people. Our findings indicate that cooperation nourishes this link, but selfish behaviour that is detrimental to others effectively compromises this link (at least with males), such that empathic responses in the brain are diminished or abolished.

These findings complement those of a previous imaging study that reported enhanced activation in dorsal striatum (caudate nucleus) when individuals punished defectors (by delivering ‘punishment points’) in a sequential Prisoner's Dilemma game²¹. In the present study we observed activation correlated with revenge in ventral striatum. This difference in evoked activity in dorsal and ventral regions of the striatum is likely to reflect the different nature of the tasks used. In the previous study²¹ subjects were required to select an action to administer punishment, whereas in the present study subjects passively observed a cue indicating that a defector was receiving pain. These findings are consistent with the different functions associated with distinct regions of the striatum: afferent projections to dorsal striatum are thought to be crucial for learning correct actions so as to maximize reward, whereas projections to ventral striatum, including nucleus accumbens, have a key function in reward prediction and pavlovian learning^{18-20,23,24}. The findings of enhanced activation in ventral striatum to a signal indicating that a defector is receiving pain are in agreement with the hypothesis that humans derive satisfaction simply from seeing justice administered^{15,21}, even if the instrument of punishment is out of their control.

Our results suggest a neural foundation for theories of social preferences. These theories^{4,7} suggest that people value the gains of others positively if they are perceived to behave fairly, but value others' gains negatively if they behave unfairly. This pattern of preferences implies that people like cooperating with fair opponents but also like punishing unfair opponents. Our corresponding neurobiological observations were more prominent in men, although further experiments are needed to confirm the gender specificity of the effect. It is possible that our

experimental design favoured men because the modality of punishment was related to physical threat, as opposed to psychological or financial threat. Alternatively, these findings could indicate a predominant role for males in the maintenance of justice and punishment of norm violation in human societies.

METHODS

A full description of experimental design and methods is provided in Supplementary Information, and an overview of the design is provided in Supplementary Fig. 1.

Subjects and confederates

Subjects (16 men and 16 women) gave informed consent and the study was approved by the Joint Ethics Committee of the National Hospital for Neurology and Neurosurgery (University College London Hospitals NHS Foundation Trust) and Institute of Neurology (University College London). We employed four professional actors (two female and two male) as confederates.

Induction of liking and disliking

Subjects played a sequential iterated Prisoner's Dilemma game, in which a first player can trust a second player by sending 10 starting points (subsequently converted to money) to the other player knowing that each point sent will be tripled²⁵. The second player (confederate) then reciprocates by sending an amount between 0 and 10 points back, which is also tripled. In each game, two confederates, sitting in different rooms, were playing fairly or unfairly by reciprocating large or small amounts of money.

Empathy for pain model

We used an electrical pain stimulus applied to the dorsum of the right hand¹⁰. Before the experiment we determined individual current amplitudes for high and low intensity stimulation, for the subject and both confederates (see Supplementary Table 1). The two confederates sat either side of the subject (positioned in the scanner) allowing the subject, with the aid of mirrors, to see all three hands. Each 12-s trial involved presentation of a visual cue, which was followed after 6s by a small circle (for 2 s) indicating the beginning of electrical stimulation (Fig. 1c). The cue indicated whether the subject (self), the fair player or the unfair player would get low stimulation (no pain) or high stimulation (pain). Stimulation intensity was indicated by the colour intensity of the cue. Each empathy for pain session consisted often trials of each of the six conditions (pain and no-pain in the context of self, fair and unfair) and 20 null events.

Although the two confederates were always of the same sex, all four possible gender combinations between the sex of the subject and the sex of the confederates were used equally often throughout the study. The position (left or right) of the actor (fair or unfair) was counterbalanced across subjects.

Post-scan questionnaires

After scanning, subjects completed a standard empathy scale²² and were asked to rate the intensity of the low and high stimulation, their liking for the two confederates, and their desire for revenge on the two confederates (see Fig. 1b, Supplementary Table 1 and Supplementary Fig. 3).

Image acquisition and analysis

We used a 1.5-T Siemens Sonata MRI scanner to acquire gradient-echo, T*2-weighted echo-planar images with blood-oxygenation-level-dependent contrast. An additional T1-weighted structural image was acquired for each subject.

After correction for head movements and spatial normalization²⁶, the images were analysed with SPM2 (Wellcome Department of Imaging Neuroscience, London) using an event-related model²⁷. The experiment constituted a 2 × 3 × 2 factorial design with the first factor representing ‘intensity of stimulation’ (pain versus no pain), the second factor ‘addressee’ (self, fair and unfair) and the third factor ‘gender’ (male or female subject).

To create regressors of interest, each condition was modelled by convolving delta functions at each trial onset (presentation of the anticipatory cue) and at each pain onset (presentation of the circle) with a canonical haemodynamic response function. Contrast images were calculated by applying linear contrasts to the parameter estimates for the regressor of each event. The contrast images were then entered into one-sample t-tests, separately for female and male subjects, to instantiate random-effects group analyses^{28,29}.

To assess shared networks of pain-related activation in self, fair and unfair conditions, we performed a conjunction analysis and an additional (more conservative) inclusive masking procedure in which we masked the pain–no-pain contrast in the fair or unfair condition with the pain–no-pain contrast in self.

Finally, we used a regression analysis to explore which brain regions showed, first, a correlation between empathy-related activity (pain–no pain in fair) and individual empathic character traits as assessed by post-scan empathy questionnaires, and second, a correlation between pain-related activity in the unfair condition compared with the fair condition (that is, pain in unfair–pain in fair) and individual tendencies to seek revenge as assessed by subjective rating scales.

On the basis of previous findings^{10–12} we expected two regions to be of particular importance for pain-related empathy: FI and ACC. In the analysis focusing on activity specific to perceiving pain in unfair players in comparison with fair players we extended our regions of interest to include areas known to be involved in reward processing including ventral striatum (nucleus accumbens) and orbito-frontal cortex. We report results at $P < 0.005$ uncorrected for multiple comparisons in the a priori regions of interest.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank P. Aston, S. Kiebel and E. Featherstone for their help. This work was supported by the German Academy of Natural Sciences Leopoldina, Halle, from the Ministry of Education and Science, by the Medical Research Council (UK), and by Wellcome Trust Programme Grants to R.J.D. and C.D.F.

References

1. Preston SD, de-Waal FBM. Empathy: Its ultimate and proximate bases. *Behav. Brain Sci* 2002;25:1–72. [PubMed: 12625087]
2. Gallese V, Keysers C, Rizzolatti G. A unifying view of the basis of social cognition. *Trends Cogn. Sci* 2004;8:396–403. [PubMed: 15350240]
3. Decety J, Jackson PL. The functional architecture of human empathy. *Behav. Cogn. Neurosci. Rev* 2004;3:71–100. [PubMed: 15537986]

4. Rabin M. Incorporating fairness into game theory and economics. *Am. Econ. Rev* 1993;83:1281–1302.
5. Levine DK. Modeling altruism and spitefulness in experiments. *Rev. Econ. Dynam* 1998;1:593–622.
6. Camerer, CF. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton Univ. Press; Princeton: 2003.
7. Fehr E, Schmidt KM. A theory of fairness, competition, and cooperation. *Q. J. Econ* 1999;114:817–868.
8. Wicker B, et al. Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust. *Neuron* 2003;40:655–664. [PubMed: 14642287]
9. Keysers C, et al. A touching sight: SII/PV activation during the observation and experience of touch. *Neuron* 2004;42:335–346. [PubMed: 15091347]
10. Singer T, et al. Empathy for pain involves the affective but not sensory components of pain. *Science* 2004;303:1157–1162. [PubMed: 14976305]
11. Jackson PL, Meltzoff AN, Decety J. How do we perceive the pain of others: A window into the neural processes involved in empathy. *Neuroimage* 2005;3:771–779. [PubMed: 15652312]
12. Morrison I, Lloyd D, di Pellegrino G, Roberts N. Vicarious responses to pain in anterior cingulate cortex: is empathy a multisensory issue? *Cogn. Affect. Behav. Neurosci* 2004;4:270–278. [PubMed: 15460933]
13. Craig AD. Human feelings: why are some more aware than others? *Trends Cogn. Sci* 2004;8:239–241. [PubMed: 15165543]
14. Singer T, Kiebel SJ, Winston JS, Dolan RJ, Frith CD. Brain responses to the acquired moral status of faces. *Neuron* 2004;41:653–662. [PubMed: 14980212]
15. Fehr E, Gächter S. Altruistic punishment in humans. *Nature* 2002;415:137–140. [PubMed: 11805825]
16. Fehr E, Fischbacher U. The nature of human altruism. *Nature* 2003;425:785–791. [PubMed: 14574401]
17. Boyd R, Gintis H, Bowles S, Richerson PJ. The evolution of altruistic punishment. *Proc. Natl Acad. Sci. USA* 2003;100:3531–3535. [PubMed: 12631700]
18. Cardinal RN, Parkinson JA, Hall J, Everitt BJ. Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neurosci. Biobehav. Rev* 2002;26:321–352. [PubMed: 12034134]
19. O'Doherty J, et al. Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* 2004;304:452–454. [PubMed: 15087550]
20. Schultz W. Multiple reward signals in the brain. *Nature Rev. Neurosci* 2000;1:199–207. [PubMed: 11257908]
21. de Quervain DJ, et al. The neural basis of altruistic punishment. *Science* 2004;305:1254–1258. [PubMed: 15333831]
22. Davis MA. A multidimensional approach to individual differences in empathy. *JSAS Cat. Selected Docs Psychol* 1980;10:85.
23. Dayan P, Balleine BW. Reward, motivation, and reinforcement learning. *Neuron* 2002;36:285–298. [PubMed: 12383782]
24. Montague PR, Hyman SE, Cohen JD. Computational roles for dopamine in behavioural control. *Nature* 2004;431:760–767. [PubMed: 15483596]
25. Berg J, Dickhaut J, McCabe K. Trust, reciprocity and social history. *Games Econ. Behav* 1995;10:122–142.
26. Friston KJ, et al. Spatial registration and normalization of images. *Hum. Brain Mapp* 1995;2:165–189.
27. Josephs O, Turner R, Friston K. Event-related fMRI. *Hum. Brain Mapp* 1997;5:243–248.
28. Friston KJ, et al. Statistical parametric maps in functional imaging: A general linear approach. *Hum. Brain Mapp* 1995;2:189–210.
29. Penny, W.; Holmes, AP. *Human Brain Function*. Frackowiak, RSJ., et al., editors. Elsevier; San Diego: 2004. p. 843-850.

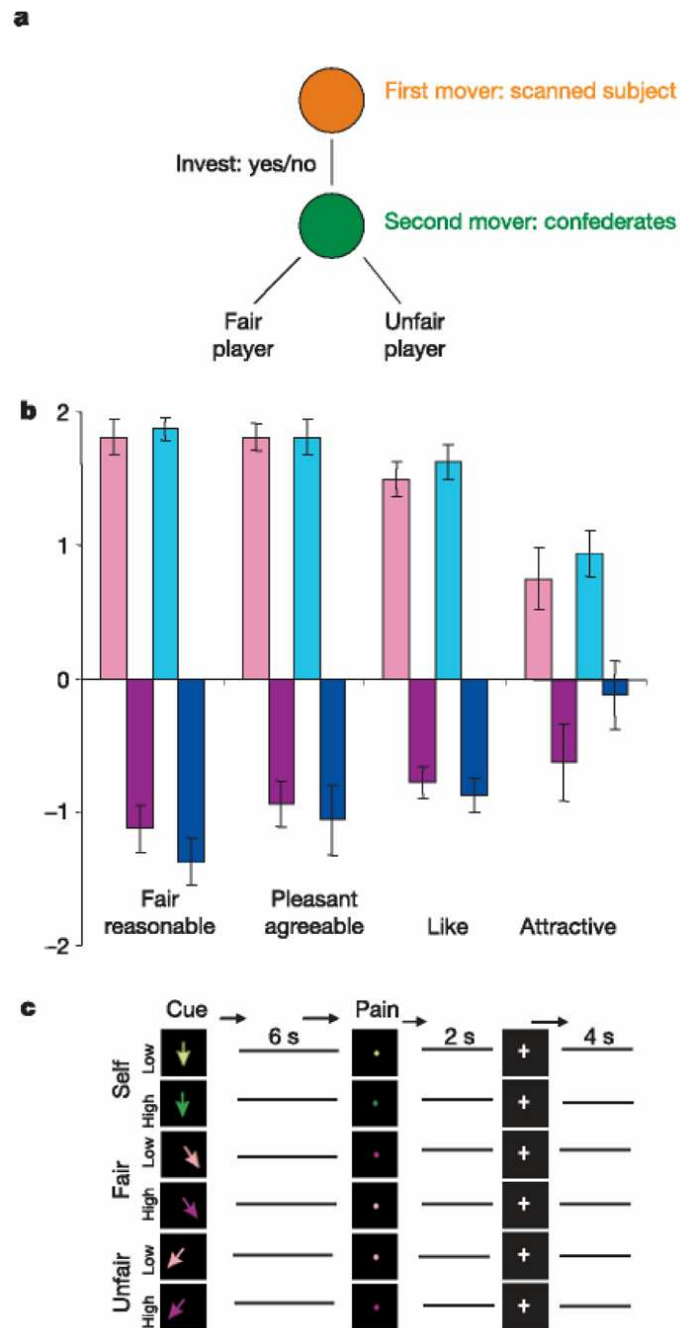


Figure 1.

Experimental models I and II and behavioural ratings. **a**, Game tree of a sequential Prisoner's Dilemma game used to induce liking and disliking of opponent players. First movers (subjects) chose between keeping and sending money to the second mover. One of the confederates sent fair (large) amounts of money back, the other unfair (small) amounts. **b**, Post-scan ratings showing that both sexes rated the fair confederates as being more reasonable/fair, their personality as being more agreeable/pleasant, as being more likeable and more attractive than unfair players (scale ranging from 22 to p2). Pink bars, women subjects rating the fair player; purple bars, women subjects rating the unfair player; cyan bars, men subjects rating the fair player; blue bars, men subjects rating the unfair player. Error bars refer to s.e.m. **c**, Experimental

design of the 'empathy for pain model'. Arrows of different colours indicate painful or non-painful stimulation applied to the scanned subject (self), the fair or the unfair player.

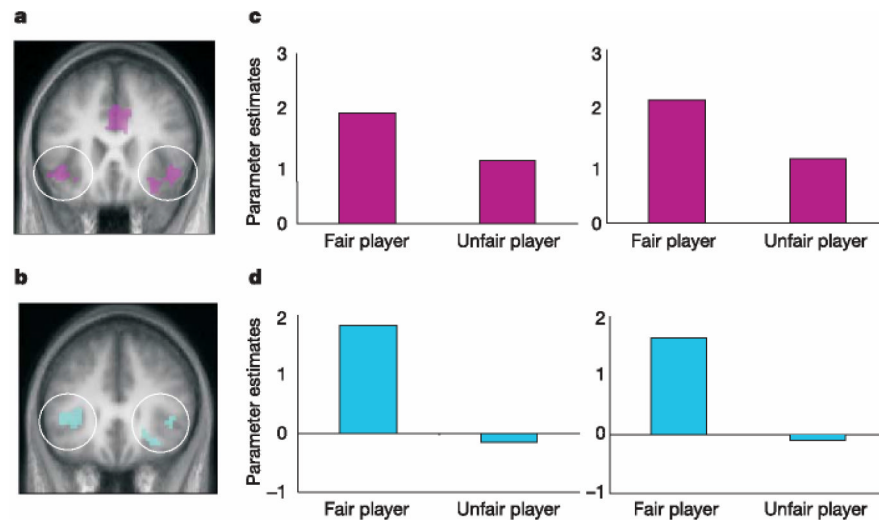


Figure 2.

Pain-sensitive activation networks to the sight of fair and unfair players in pain. a, b, Conjunction analysis between the contrasts pain–no pain in the context of self and the fair condition at $P < 0.001$ for women (pink, a) and men (blue, b). Increased pain-related activation (asterisk indicates whole-brain corrected) for women in ACC* [9,18,27], left FI* [−42,15,−3], right FI* [30,18,−18], left SII* [−60,−30,18], right SII* [63,−30,24] and brainstem* [3,−18,−18]; for men in left FI* [−33,33,3] right FI [42,33,3] and brainstem [3,−33,−30]. c, d, Average activation (parameter estimates) in peak voxels of left and right FI (left and right panels, respectively) for the painful–non painful trials in fair and unfair conditions for women (c) and men (d).

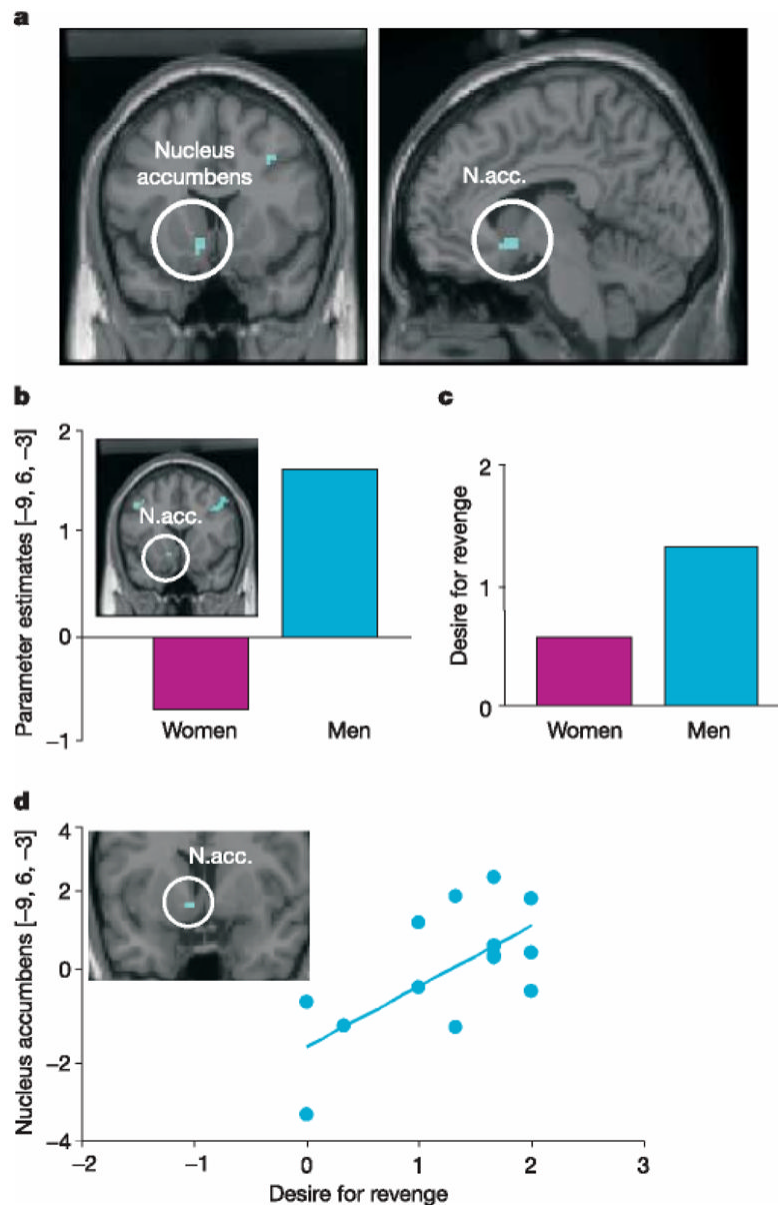


Figure 3.

Gender differences in brain activity in nucleus accumbens specific to the perception of an unfair compared to fair player in pain. a, Increased activity ($P < 0.005$) in nucleus accumbens $[-9, 15, -9]$ for painful trials in the unfair–fair condition for men but not for women. b, Average activation (parameter estimates) for women (pink) and men (blue) in left nucleus accumbens $[-9, 15, -9]$ when testing for gender differences. c, Men (blue) indicate stronger feelings of desire for revenge than women (pink) ($t(30) = 2.40$, $P < 0.05$) measured on a scale from -2 ('not at all') to 2 ('very much'). d, Correlation ($r = 0.68$, $P < 0.05$) of parameter estimates at peak of nucleus accumbens activation $[-9, 6, -3]$ for the (pain in unfair–pain in fair) contrast in men with expressed desire for revenge in men. There was no correlation for women.